**PART A:**

**Q1 Run a linear regression on data provided, do mention list of things mentioned below after running**

Dependent Variable

Independent Variables

Number of Columns

Number of Rows

Categorical Variables

Continuous Variables

Significant Variables

Beta Coefficients

Regression Equation

Interpretation of equation

**Q2 Run a logistic regression on data provided, do mention list of things mentioned below after running**

Dependent Variable

Independent Variables

Number of Columns

Number of Rows

Categorical Variables

Continuous Variables

Significant Variables

Beta Coefficients

Equation

Interpretation of equation

F1 Score, Accuracy, Precision, Recall, Confusion matrix

**Q3 Run a CART tree on data provided, do mention list of things mentioned below after running**

Dependent Variable

Independent Variables

Number of Columns

Number of Rows

Categorical Variables

Continuous Variables

Interpretation of parent node (first 3 nodes)

F1 Score, Accuracy, Precision, Recall, Confusion matrix

**Q4 Run a KMEANS clustering on data provided, do mention list of things mentioned below after running**

Number of Columns

Number of Rows

Categorical Variables

Continuous Variables

Number of significant K

Silhouette Score for Kmeans model

**PART B**

**Q1 Run a linear regression on data provided, do mention list of things mentioned below after running**

**Run a model with only set of significant variables**

Dependent Variables

Independent Variables

Number of Columns

Number of Rows

Categorical Variables

Continuous Variables

Variables effected with outliers

Variables effected with missing value

Significant Variables

Beta Coefficients

Regression Equation

Interpretation of equation

MAPE for Test & Train

**Q2 Run a logistic regression on data provided, do mention list of things mentioned below after running. Run a final model with only significant variables**

Dependent Variables

Independent Variables

Number of Columns

Number of Rows

Categorical Variables

Continuous Variables

Variables effected with outliers

Variables effected with missing value

Significant Variables

Beta Coefficients

Regression Equation

Interpretation of equation

Set a cuttoff value to define a class by optimal value from ROC

F1 Score, Accuracy, Precision, Recall, Confusion matrix for both test & train

**Q3 Run a CART and CHAID tree on data provided, do mention list of things mentioned below after running. (repeat the steps for both CART & CHAID)**

Dependent Variables

Independent Variables

Number of Columns

Number of Rows

Categorical Variables

Continuous Variables

Interpret first 3 nodes

F1 Score, Accuracy, Precision, Recall, Confusion matrix for both test & train

**Q4 Run a KMEANS clustering on data provided, do mention list of things mentioned below after running. (Develop 4 different models and select the best model)**

Number of Columns

Number of Rows

Categorical Variables

Continuous Variables

Number of significant K

Silhouette Score for Kmeans model